

Analysis of heat kernel highlights the strongly modular and heat-preserving structure of proteins

Lorenzo Livi^{*†1}, Enrico Maiorino^{‡2}, Andrea Pinna^{§3}, Alireza Sadeghian^{¶1}, Antonello Rizzi^{||2}, and Alessandro Giuliani^{**4}

¹Dept. of Computer Science, Ryerson University, 350 Victoria Street, Toronto, ON M5B 2K3, Canada

²Dept. of Information Engineering, Electronics, and Telecommunications, SAPIENZA University of Rome, Via Eudossiana 18, 00184 Rome, Italy

³CRS4 Bioinformatica, Parco Tecnologico della Sardegna, Edificio 1, 09010 Pula (CA), Italy

⁴Dept. of Environment and Health, Istituto Superiore di Sanità, Viale Regina Elena 299, 00161 Rome, Italy

March 17, 2015

Abstract

In this paper, we study the structure and dynamical properties of protein contact networks with respect to other biological networks, together with simulated archetypal models acting as probes. We consider both classical topological descriptors, such as the modularity and statistics of the shortest paths, and different interpretations in terms of diffusion provided by the discrete heat kernel, which is elaborated from the normalized graph Laplacians. A principal component analysis shows high discrimination among the network types, either by considering the topological and heat kernel based vector characterizations. Furthermore, a canonical correlation analysis demonstrates the strong agreement among those two characterizations, providing thus an important justification in terms of interpretability for the heat kernel. Finally, and most importantly, the focused analysis of the heat kernel provides a way to yield insights on the fact that proteins have to satisfy specific structural design constraints that the other considered networks do not need to obey. Notably, the heat trace decay of an ensemble of varying-size proteins denotes subdiffusion, a peculiar property of proteins.

Keywords— Protein contact networks; Heat kernel; Subdiffusion; Network analysis.

1 Introduction

As aptly pointed out by Nicosia et al. [61] “Networks are the fabric of complex systems”. Concepts like organized complexity (or middle way [46], mesoscopic systems [43]) revolve around the existence of features shared by systems made of many interacting parts, which can be suitably described in terms of networks.

^{*}llivi@scs.ryerson.ca

[†]Corresponding author

[‡]enrico.maiorino@uniroma1.it

[§]andrea.pinna@crs4.it

[¶]asadeghi@ryerson.ca

^{||}antonello.rizzi@uniroma1.it

^{**}alessandro.giuliani@iss.it

In 1948 Weaver [78] defined the notions of simplicity and complexity in Science by a three categories classification: (1) *Organized Simplicity*. The paradigm of this kind of problems is exemplified by Newton’s law of universal gravitation; (2) *Disorganized Complexity*. In these cases, systems are made by a very large number of interacting elements. Even if these systems cannot be studied at the single-element level, nevertheless they allow for a very efficient statistical treatment (e.g., in terms of statistical mechanics). There is a third kind of complexity that cannot be studied this way and that Weaver identified with biological systems: *Organized Complexity*.

Both the arguments and terminology used by Weaver were largely coincident with Laughlin *et al.*, identifying “The Middle Way” [46] as the real frontier of basic Science in the XXI century. The natural entities in which the organized complexity approach was more fruitful were protein molecules. Not only they were exactly at the boundary between simple and complex physics [31] but they allowed for both very accurate experiments and for a very refined structural description (e.g., via X-ray crystallography). The situation nowadays continues to be unchanged with respect to Laughlin *et al.* [46]: protein science is still the most explored field in the realm of organized complexity [4, 5, 15, 24, 63, 77, 80].

The resurgent interest in graph-theoretical methods for the analysis complex systems [9, 11, 14, 18–20, 22, 24, 26, 34, 42, 52, 53, 58, 60, 70, 72, 73] set forth by the works by Barabasi, Strogatz and other pioneers in the first decade of XXI century, allowed for a rich set of measures providing a multifaceted description of complex networks [3, 27, 29, 29, 35, 40, 52, 58, 68, 69]. It is sufficient that a given problem is formalized by means of a vertex-and-edge representation – with vertices being the parts and edges their pairwise relations – to allow facing the problem according with the terms of organized complexity [34]. Protein molecules are at the forefront of complex network way to organized complexity. In fact, it is possible to represent a protein molecule in terms of protein contact network (i.e., a network having amino acid residues as vertices and edges representing physical contacts between them) by considering its native 3D structure [30, 80]. It is worth noting that the graph-based representation is at the core of chemical thought: the structural formulas are graphs [75].

In this work, we face a data-driven quest for mesoscopic organization principles of complex biological systems by analyzing different complex networks: protein contact networks, metabolic networks, and genetic networks, together with simulated archetypal networks whose wiring scheme is generated by mathematical rules. All considered networks are characterized in terms of two collections of numerical features. The first collection is based on classical topological descriptors, such as the modularity and statistics of the shortest paths. The second one exploits the discrete heat kernel (HK), elaborated using the eigendecomposition of the (normalized) graph Laplacian [79]. With a first preliminary analysis, we show that the different classes of networks are discriminated by a suitable embedding of the numeric features. This is reasonably expected given the substantially different natures of the analyzed networks, but by no means can be considered as a trivial result. As a matter of fact, the distinction in terms of metabolic, genetic, and protein contact networks is based on network functions, and the demonstration of a link between functional and structural properties of the corresponding graph representation is a prerequisite for the soundness of the proposed strategy of analysis. An important result is that the two herein considered network characterizations resulted to be strongly correlated with each other, so giving a proof-of-concept of the reliability and interpretability of the adopted network descriptions. From this first analysis, it also emerged that protein contact networks display unique properties that do not allow for a straightforward classification in any of the considered classical archetypal networks, stressing the need for the search of new generative models for protein contact networks. The second, and more important, claim of this paper is that the ensemble HK analysis allows us to derive results in agreement with known chemico-physical properties of protein molecules. Our analysis demonstrates that a (simulated) diffusion process in protein contact networks proceeds slower than normal diffusion, i.e., we observe subdiffusion. Notably, a two-regime diffusion emerged from the analysis of the heat trace decay: a fast and a slow regime. The fast regime is driven by short cuts putting in contact amino acid residues far-away along the sequence. Subdiffusion is ubiquitous in Nature [41] and it is a well-studied peculiarity describing energy flow [47–49, 71] and vibration dynamics [36, 57, 66, 67, 81] in protein molecules. Interestingly, here we are able to observe the same phenomenon by exploiting only algebraic properties of the considered graph representations of an ensemble of varying-size proteins. The demonstration of the ability

of such a network formalization to explain a unique chemico-physical property of protein molecules, as heat diffusion, is a proof-of-concept of the usefulness of graph theory in chemical systems modeling.

There is sufficient agreement on the fact that proteins, considering their native structure, are highly modular and fractal networks [6, 23, 25, 28, 52, 55]; yet they are characterized also by short paths connecting distant regions of the molecules responsible for the fast-track transport of energy and protein allosteric properties [47]. The modular character of protein contact networks has a direct impact on diffusion processes simulated via graph-theoretical methods. In fact, well-defined modular organizations slow down diffusion processes in networks [32]. Additionally, theoretical and experimental results regarding diffusion on porous and fractal media predict fractional scaling exponents for the mean squared displacement (a well-known measure of diffusion), which give rise to the so-called anomalous diffusion [12, 56]. To conclude, another interesting fact deduced from our analysis is that, at odds with the other networks, modularity of protein graphs increases with the size of the network. This observation is fairly interesting since path efficiency and modular properties are two conflicting features in networks [16, 33], which however seem to be suitably optimized in proteins.

The remainder of this paper is organized as follows. In Sec. 2 we first introduce the data that we considered in our study (Sec. 2.1) and successively we describe the two adopted network characterizations (Sec. 2.2). In Sec. 3 we present the experimental results. In Sec. 3.1 we discuss the results in terms Principal Component Analysis (PCA) of the considered network characterizations in terms of numerical features; in Sec. 3.2 we demonstrate the important statistical agreement among the two characterizations. Finally, in Sec. 3.3 we present the principal results of this paper, in which we show the two-regime diffusion of proteins. Sec. 4 concludes the paper; Appendix A provides the HK technical details and Appendix B offers a justification for considering ensemble statistics in the analysis of the HT decay.

2 Material and Methods

2.1 The Considered Dataset

Our dataset consists of 323 connected networks (simple graphs). We considered 100 randomly selected E.Coli protein contact networks (PCN) from the dataset recently elaborated by us [51, 52] – in the literature PCN are also referred to as protein contact maps, although here we consider the two denominations as equivalent. Such proteins have been obtained by integrating the Niwa *et al.* [62] E.Coli data with the available information of the respective native structures gathered from the Protein Data Bank repository [2]. The selected proteins contain from 300 to 1000 amino acid residues. As prescribed by Di Paola *et al.* [24], undirected edges are added among any two alpha carbon atoms within the $[4, 8]$ Å range; the lower 4 Å filter allows to get rid of trivial contacts that do not have important topological properties. Then, we considered 43 metabolic networks (MN) describing organisms belonging to all three domains of life. Vertices of such networks are the substrates and product of the chemical reactions, while the edges are the reactions catalyzed by the enzymes. As demonstrated in Ref. [44], those large metabolic networks exhibit a typical scale-free topology. The size of the 43 metabolic networks ranges from 300 to 1500 vertices. We also considered a collection of 50 realistic gene regulatory networks (GEN) with a number of vertices varying from 200 to 1100 genes/vertices. The GEN are generated with the SysGenSIM software [65], a MATLABTM toolbox for the simulation of systems genetics datasets. Artificial networks and data by SysGenSIM have already been officially employed for the verification of gene network inference algorithms, such as in the DREAM5 Systems Genetics challenge [1]; they have also been used as benchmarks for the development of state-of-the-art reverse-engineering algorithms [64]. The considered GEN networks have been generated with the Exponential Input Power-law Output (EIPO) model, i.e., they are built by (i) sampling the number of ingoing and outgoing edges for each vertex from, respectively, an exponential and a power law distribution, and then by (ii) connecting the vertices accordingly. These artificial EIPO networks exhibit two well-known structural characteristics of real gene networks: the modularity [8], and the vertex in-degree and out-degree distributions fitting, respectively, an exponential and a power-law curve [37]. Besides the adopted EIPO topology, we considered an average vertex degree varying from 4 to 8: the average degree has been sampled

in such a wide range due to the uncertainty in the size of the interactome in typical gene regulatory networks. Apparently, the complexity of biological organisms better correlates with the number of interactions between genes than with the number of genes. Therefore, the average number of edges in gene regulatory networks varies according to the complexity of the represented organism [74]; it makes then sense to study gene regulatory networks with a different number of interactions/edges.

To obtain suitable references with the aim of helping us in discussing the results, we considered 130 additional networks of varying size belonging to well-known classes of graphs – such networks play the role of “probes” in our dataset. Notably, we considered 10 Erdős-Rényi (ER) graphs generated with probability $p = \log(n)/n$; 10 Barabási-Albert (BA) scale-free networks [7] with a six-edges preferential attachment scheme; and 10 random regular graphs (REG) with degree equal to six. To cover all network sizes, such probe networks are generated with a number of vertices ranging from 200 to 1100. Finally, we also considered the synthetic counterpart of the 100 real proteins (denoted as PCN-S in the following). Such synthetic proteins have been generated by considering the same number of vertices and edges of the real proteins. The generation mechanism of the topology follows the three-rule scheme proposed by Bartoli et al. [10], to simulate the folded configuration of the protein backbone by a probability of contact decreasing with the sequence distance. The only exception is for the rule involving edges of the backbone structure. In fact, to be consistent with the architecture of our real proteins (we considered edges among residues within 4–8 Å), in PCN-S we added edges only among consecutive residues in the sequence having distance 2. It is worth pointing out that such a generation mechanism gives rise to networks with typical small world topology [10].

2.2 Characterization of the Graph Topology

In the following, we provide an essential description of the two graph characterizations used in this paper. Full details are omitted for the sake of brevity; we include references to the literature. In practice, each characterization is meant to offer a description of the original graphs as vectors of numerical features. The first characterization employs “classical” topological descriptors (TD), which include statistics of the degrees/shortest paths and also elaborations of the graph spectrum. In particular we consider the number of vertices (V) and edges (E) as basic descriptors of the size of the network; the modularity (MOD) [13, 59] for quantifying the presence of a global community/cluster structure (please note that we consider as feature the value associated to the partition with maximum modularity); the average closeness centrality (ACC), average shortest path (ASP), average degree centrality (ADC), and average clustering coefficient (ACL) [17]; the energy (EN) and Laplacian energy (LEN) of the spectrum [39]; two invariant features from the heat kernel – see later for details; the ambiguity (A) [50], which expresses the degree of irregularity of the topology; and finally the entropy of a stationary Markovian random walk (H) [21].

In the second characterization we exploit the HK only, whose technical details are reported in Appendix A. In this respect, we consider here three sets of features: those extracted from the heat trace (HT), the heat content (HC), and the series of invariants (8) associated to the HC, which are called heat content invariants (HCI). Please note that HT and HC are time-dependent characteristics, while HCI is not. Therefore, in the characterization exploiting the HK only, we consider several time instants for HT and HC; in the TD characterization, instead, we consider only one time instant for the HT – corresponding to a “transient” regime of the diffusion – and only the first HCI coefficient. Further details are progressively provided in the following sections.

3 Results

The results of the computational experiments are organized in the following subsections. In Sec. 3.1 we show and discuss the PCA performed over the two aforementioned characterizations of the considered networks. In Sec. 3.2 we discuss the canonical correlation analysis (CCA) calculated among these two characterizations. Finally, in Sec. 3.3 we present the results obtained in terms of scaling and diffusion dynamics.

3.1 Principal Component Analysis

Numeric data have been standardized using the component-wise mean and standard deviation. TD and HK fields were submitted to two separated PCA, projecting hence the networks in the spaces spanned by respective principal components (PC). The emerging of local models specific for each type of network, and the correlation of the components extracted on the entire data set, is thus a proof of the presence of different architectures characterizing the considered classes of networks. Moreover, the mutual correlation between PCs of TD and HK – assessed in Sec. 3.2 by a CCA – gave us a demonstration of the consistency and interpretability of the considered network descriptions.

3.1.1 Analysis of Topological Descriptors

Fig. 1 shows the PCA of the topological descriptors (PCA-TD). The first three PCs are sufficient to explain more than 90% of the variance ($\simeq 91\%$). As it is possible to observe in Fig. 1(a), PC1–PC2 offer a very clear discrimination among the different classes of networks. The separability persists also by considering PC1–PC3, while however we observe that GEN lose compactness and overlap with MN. By considering PC2–PC3, instead, PCN overlap with REG. However, the overall picture emerging from PCA-TD clearly points to the possibility of distinguishing among the network types.

Let us now interpret such PCs. Tab. 1 shows the loadings of the first three factors of PCA-TD. The first factor (FACTOR-1) is primarily characterized by MOD, ACC, and ASP. As MOD increases (the community structure becomes more evident) preferential paths connecting different regions of the network increase as well. In fact, ACC and ASP are, respectively, negatively and positively correlated with MOD. It is worth mentioning that ACC and ASP offer a somewhat opposite view of the same feature, i.e., the efficiency of the paths in the networks. As the global community structure emerges (captured by MOD), also the local clustering structure (ACL) increases as well, although ACL is less loaded on FACTOR-1. In addition it is worth noting the agreement among the randomness (H) and the modularity: predictability of stationary random walks is affected by the presence of network modules/communities.

FACTOR-2 positively correlates the number of vertices (V) with LEN, which clearly points to the correlation among the network size in terms of number of vertices and the global architecture. The meaning of this factor will appear more clear in Sec. 3.3, where we will discuss the scaling of the number of vertices with MOD and the invariant characteristics of the HK.

Finally, the third factor (FACTOR-3) could be interpreted as the “redundancy” of the network wiring substrate. In fact, descriptors heavily loaded on FACTOR-3 are those more directly related to the adjacency matrix–edges. The ambiguity (A) decreases as the number of edges increases. This means that adding redundancy to the network (i.e., alternative paths) affects the regularity of the topology.

It is immediate to recognize how the different types of networks are characterized by local linear models in the globally orthogonal PC spaces. These linear models correspond to different scaling relations with network size – discussed later in Sec. 3.3.

A simple look at Fig. 1 allows to catch the singular position of PCN on the extreme right of the most informative PC1–PC2 space, so confirming the peculiar character of PCN with respect to classical network architectures. Moreover it is worth noting that the artificial polymer networks – PCN-S – are the most similar to PCN, although it is not possible to appreciate any overlap. This fact suggests that proteins are not just “coiled strings” as hypothesized in Ref. [10]. In addition to the features coming from the folding of a continuous backbone, PCN have other peculiar characteristics.

3.1.2 Analysis of the Heat Kernel

We consider three types of invariant features elaborated from the HK: HT, HC, and HCI. For the PCA of HT and HC we take into account ten time instants going from $t = 0$ to $t = 9$; this choice will be justified later in Sec. 3.3; for the PCA of HCI we consider the first ten coefficients q_m of the series in Eq. 8 – this choice is motivated by the fact that for higher-order coefficients the values become numerically unstable. In all cases, the first three PCs are sufficient to explain more than 90% of the variance of the original data, and so they are retained for the embedding.

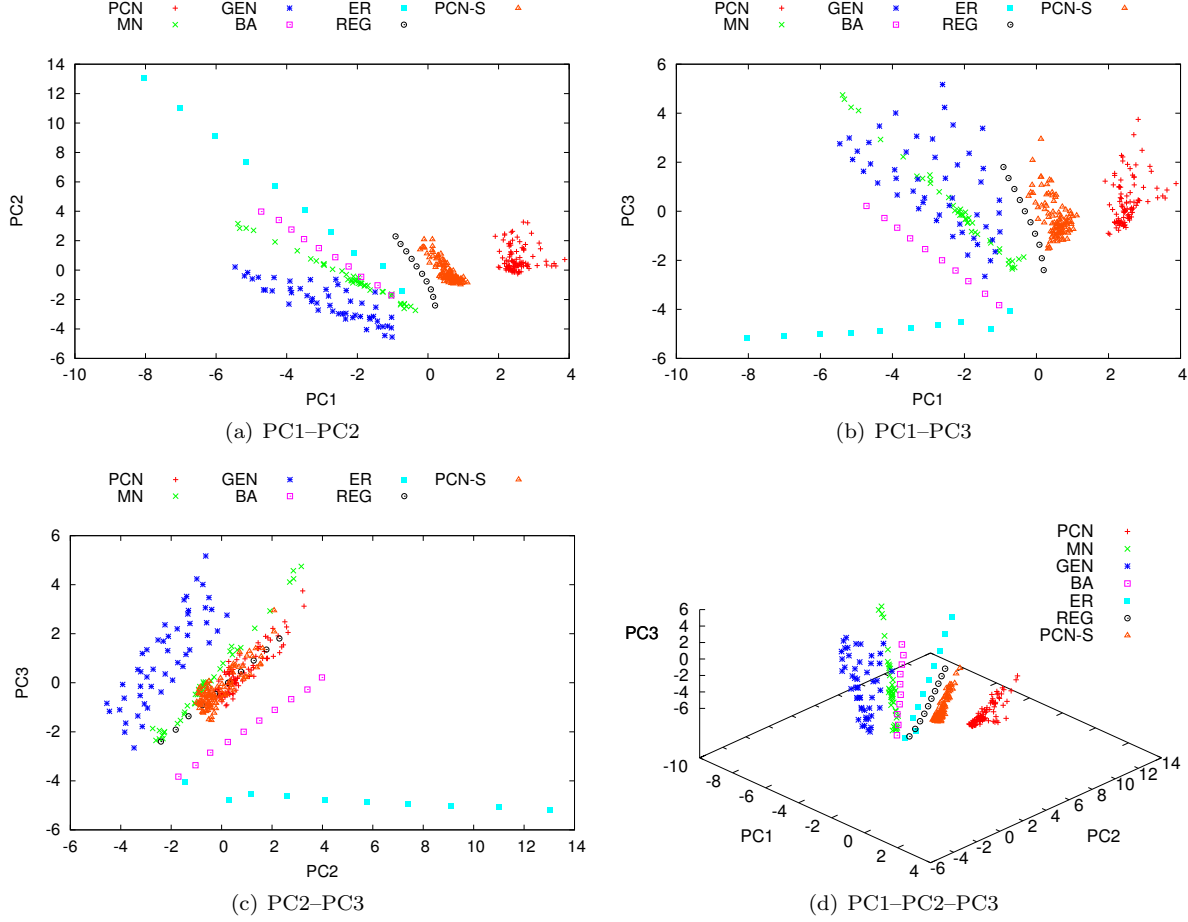


Figure 1: Embedding considering the first three PCs of PCA-TD.

Table 1: Loadings of the first three factors of PCA-TD. Relevant correlations are in bold.

DESCRIPTOR	FACTOR-1	FACTOR-2	FACTOR-3
V	-0.0441	0.9953	0.0513
E	-0.2053	0.5095	0.8082
MOD	0.9591	-0.1383	-0.1036
ADC	-0.2294	0.0428	0.9375
ACC	-0.9918	0.0353	0.0637
ASP	0.9281	-0.0279	0.0315
ACL	0.6716	-0.3588	-0.0010
EN	-0.0166	0.6830	0.7268
LEN	-0.3944	0.8407	0.1712
HT ($t = 5$)	0.6696	0.6486	-0.1007
HCI ($m = 1$)	0.4914	-0.6172	0.4639
H	0.6906	-0.2774	0.4918
A	-0.3229	0.1878	-0.7584

In Fig. 2 it is shown the PCA of the HT representation (PCA-HT). From PC1-PC2 and PC2-PC3 of PCA-HT it is possible to understand that PCN are clearly recognizable by considering the HT, while however GEN depict a not-so-coherent pattern (this is valid for all three PCs). In Fig. 3, instead, we show the PCA of the HC representation (PCA-HC). We remind to the reader that HT and HC are correlated, since HC considers the information provided by both eigenvalues and eigenvectors of the normalized Laplacian, and not just the eigenvalues as in the HT case. In fact, from PCA-HC it is possible to note that all networks

denote a clear distinguishability; considering either PC1–PC2 and PC2–PC3 almost all networks seems to denote a very peculiar configuration in the PCA space.

Finally, in Fig. 4 we show the PCA of the HCI representation (PCA-HCI). The PCs of PCA-HCI allow us noting how PCN, REG, and ER denote a very compact configuration in the PCA-HCI space, while GEN, BA, and MN present a more sparse distribution. This fact might be interpreted by observing that such two groups differentiate among networks having a clear scale-free topology (second group) and those that are not scale-free (first group). Interestingly, PCN-S seem to lie in-between those two groups.

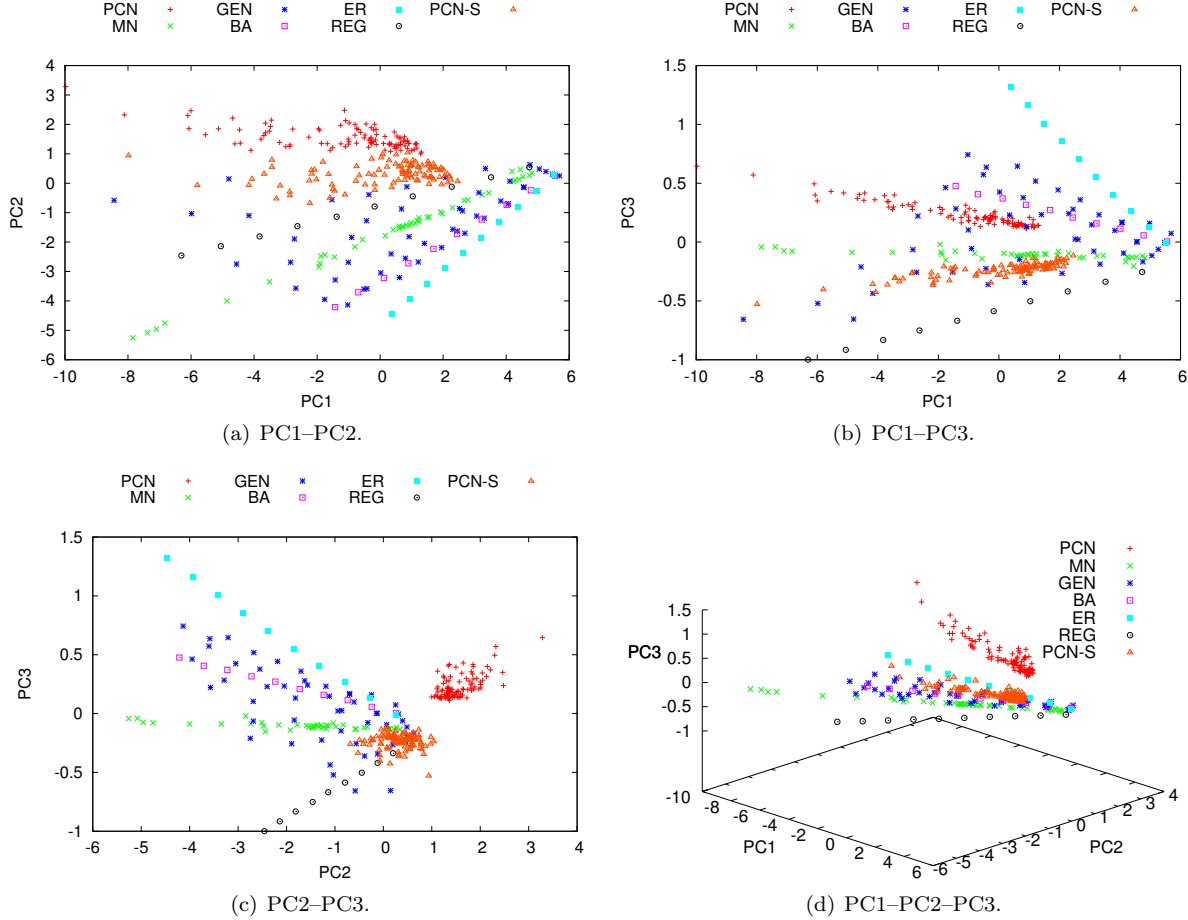


Figure 2: Embedding of the first three PCs of PCA-HT considering $t = 0, 1, \dots, 9$.

3.2 Canonical Correlation Analysis of the PCA Representations

Here we discuss the canonical correlation analysis (CCA) calculated among the various PCA discussed in Sec. 3.1. For the CCA, we always consider the first three PCs of each representation. In Tab. 2 are reported the pairwise correlation values among the most important canonical variates. There is a strong agreement among all considered PCA representations. Since part of the information from the HK is present also in TD, we have considered also a PCA representation of TD that does not include such information – in the table it is indicated as “PCA-TD_NO-HK”. Interestingly, removing the information of the HK from the TD does not alter the scored correlation, so giving a demonstration of the strong coherence between TD and HK based representations of the considered networks. This result suggests the possibility to interpret the three

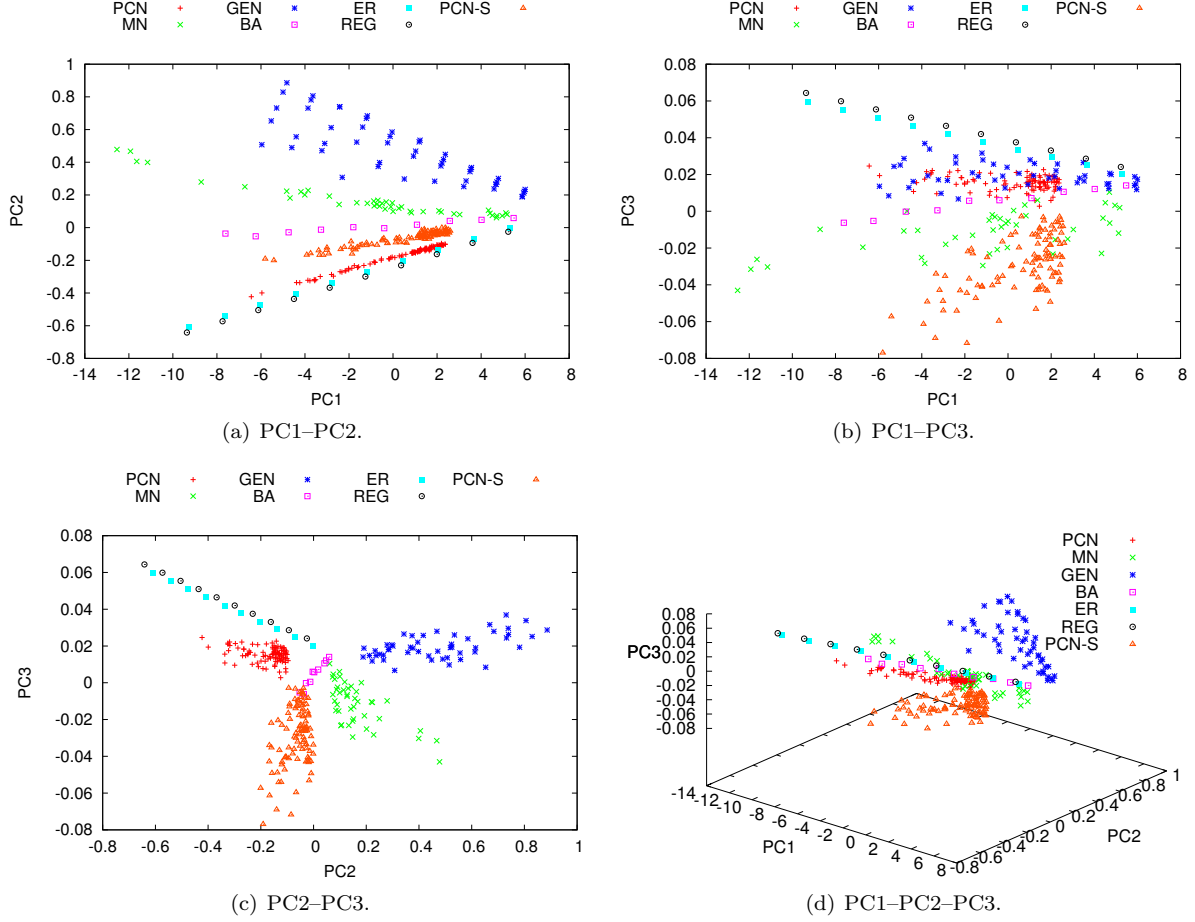


Figure 3: Embedding of the first three PCs of PCA-HC considering $t = 0, 1, \dots, 9$.

HK based representations in terms of the more understandable linear correlation structure discussed in Sec. 3.1.1.

Table 2: Canonical correlation coefficients between the first canonical variates relative to different principal component spaces.

	PCA-HT	PCA-HC	PCA-HCI
PCA-TD	0.993	0.992	0.961
PCA-TD_NO-HK	0.988	0.986	0.946

3.3 Scaling and Heat Diffusion Analysis

In this section we first study the results in terms of scaling of MOD, HT, HC, and HCI w.r.t. the number of vertices of the considered networks. Successively, we provide an analysis of the characteristic diffusion patterns emerged from the focused study of the HK.

Fig. 5 shows the scaling of the modularity with the size (number of vertices) of the networks. As already noted in Tab. 1, V and MOD do not appear to be globally correlated. In fact, PCN and PCN-S are the only architectures that show an increasing trend, while the others appear to be almost uncorrelated. We note an

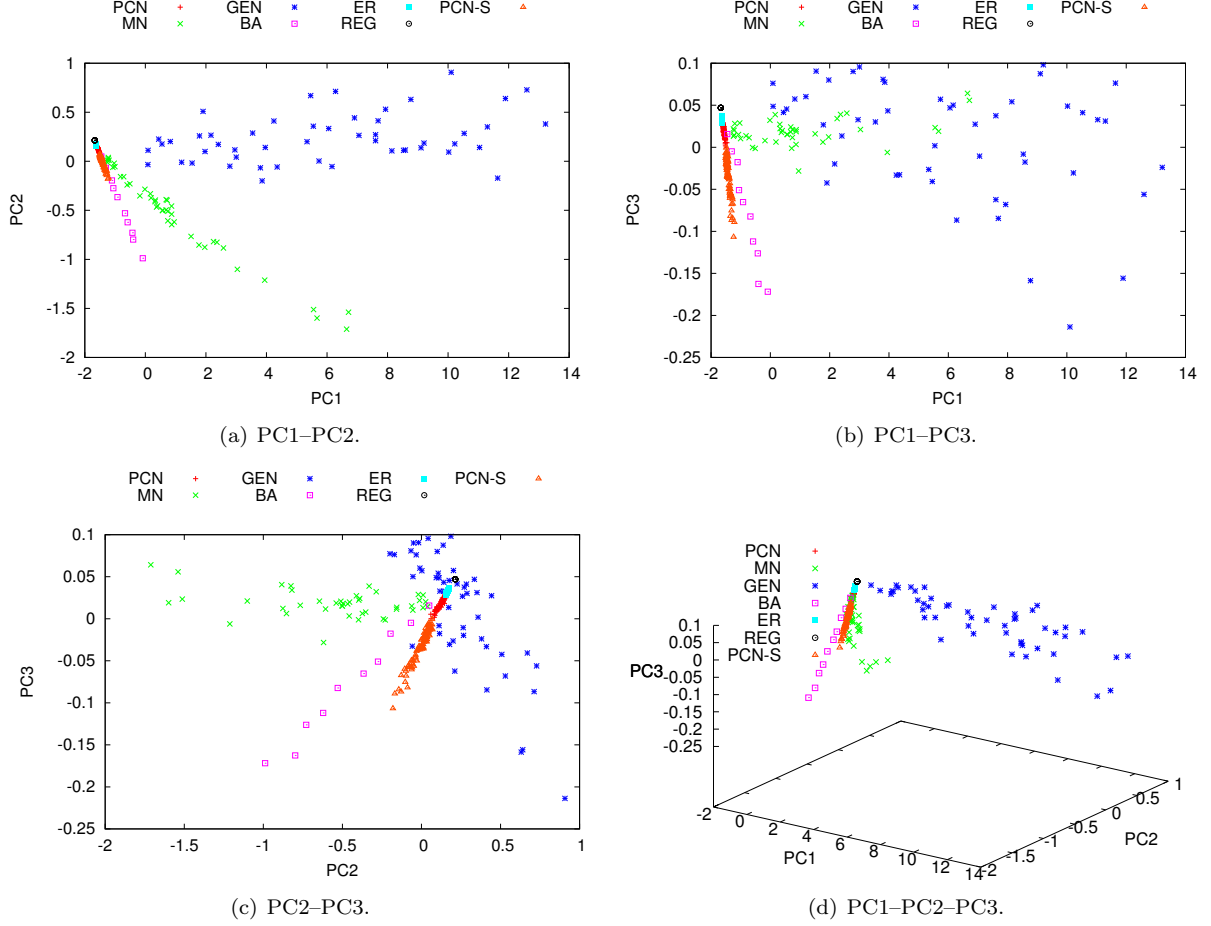


Figure 4: Embedding of the first three PCs of PCA-HCI considering the first ten HCI coefficients.

exception for ER that tend toward a negative correlation; please note that in the ER case, analytical results are available [38]. It is worth explaining the particular pattern of GEN, which does not show a clear trend. In Fig. 5(b) we show the scaling for GEN by considering the different average degrees used for the EIPO model, where we can observe that each average degree gives rise to a definite trend of MOD. Please note that the linear fitting lines in Fig. 5 are introduced only to help the reader visually, and are not meant to provide a model describing the MOD trend asymptotically.

Figs. 6, 7, and 8 show the scaling of all considered HK invariants. Initially we consider only three relevant time instants for HT, i.e., $t = 1, 5, 9$, which are depicted, respectively, in Figs. 6(a), 6(b), and 6(c). It is possible to note that, as expected, at $t = 1$ all networks show a similar increasing linear trend w.r.t. the network size. As the time instant increases, instead, PCN show a positive slope at least one order of magnitude greater than the others. At first, this fact might be attributed solely to the intrinsic high modularity characterizing the protein structures. To this end, in Fig. 6(d) we globally correlated MOD with HT over time – the time t here has a fine-grained sampling going from 0.1 to 100 with an increment step of 0.1. In the same plot, we show also the partial correlation obtained when considering the number of vertices as the control variable (indicated as “MOD-HT(t) / V” in the figure). The linear correlation trend shows that initially the two quantities are fairly anti-correlated, while they soon become very correlated, reaching the maximum correlation ($\simeq 0.88$) around the time instant $t = 10$. Successively, the correlation decreases with a smooth trend. The partial correlation demonstrates that the initial negative correlation is due to the network size effect; correlations are positive when the size is removed. This variability in the

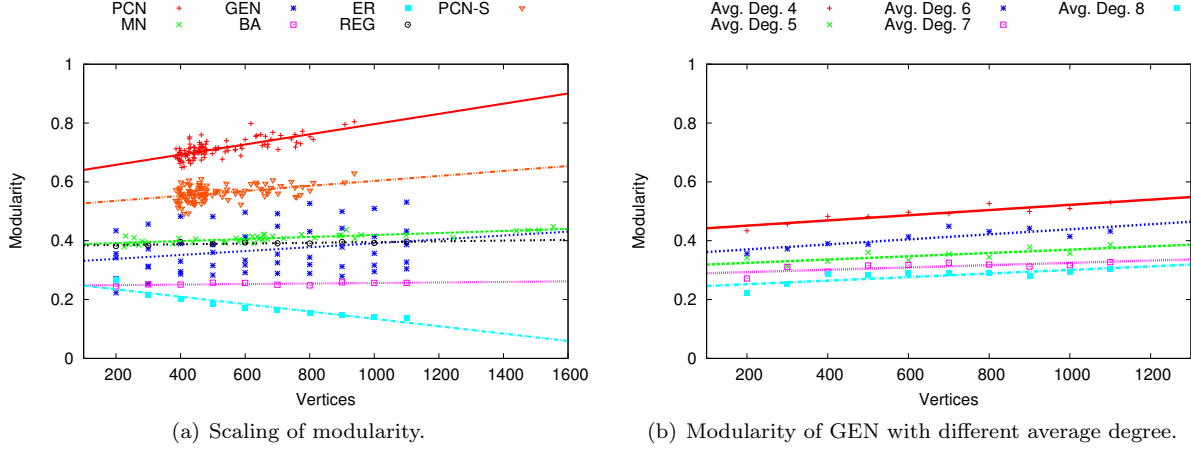


Figure 5: Scaling of modularity over network size. Please note that the linear fitting lines are introduced only to visually guide the reader, and are not meant to provide a model describing the MOD trend asymptotically.

correlation points out the fact that, although the information provided by HT is fairly consistent with the one provided by MOD, they are by no means equivalent. Notably, HT offers a richer type of information that we will further exploit in the following in terms of diffusion. In addition, this particular trend justifies the selection of the first ten time (integer-valued) instants for the calculations of PCA-HT, PCA-HC, and PCA-HCI performed in Sec. 3.1.2.

Now we elaborate over the diffusion dynamics of the considered network ensembles by analyzing the log-log plots shown in Fig. 7. Let us denote the time-dependent linear best-fitting (as those in Figs. 6(a), 6(b), and 6(c)) for the HT over the networks of a specific class, \mathcal{C} , as a function of the network size n , with the following expression:

$$\text{HT}_{\mathcal{C}}(n; t) \simeq \alpha_{\mathcal{C}}(t) \cdot n, \quad (1)$$

where n is the number of vertices and $\alpha_{\mathcal{C}}(t) \geq 0$ is the time-dependent slope. As discussed in Appendix B, by fitting linearly the HT we implicitly hypothesize the possibility to consistently describe each class of networks with an ensemble, \mathcal{C} , characterized by a unique probability density function of the normalized Laplacian eigenvalues (see Ref. [54] for a related theoretical study on a similar topic). This assumption is also justified by the results of PCA-HT reported in Fig. 2, which show good agreement among networks of the same class. As a consequence, the linear best-fitting (1) allows to consider a statistic over an entire homogeneous class of networks, instead of focusing on each isolated network dynamics separately. It is straightforward to realize that $\text{HT}_{\mathcal{C}}(n; t) = n$ for $t = 0$, i.e., $\alpha_{\mathcal{C}}(0) = 1$. As t grows, $\alpha_{\mathcal{C}}(t)$ becomes always smaller, with a rate that is related to the characteristic HT decay of the ensemble. In Fig. 7(a) we show the linear best-fitting slopes of HT, $\alpha_{\mathcal{C}}(t)$, as a function of time – note that t always varies from 0 to 100 with an increment step of 0.1. While one expects to observe trends consistent with an exponential decay (see definition of HT in Eq. 6), it is possible to recognize a different asymptotic trend for the PCN ensemble. For the sake of a better visualization, in Figs. 7(b), 7(c), and 7(d) we report the same plot but isolating, respectively, PCN, MN, and GEN; other networks are omitted for brevity. Fig. 7(b) depicts what we might consider a change of functional form for the PCN trend at some point in time (i.e., starting around $t \simeq 5$). This change of regime in the diffusion lasts few time instants, then the trend switches from exponential to power law like. This is not noted for the other networks that, instead, remain consistent with an exponential decay – they are actually expressed as sums of exponentials. In practice, from a certain $\tilde{t} > 5$, the diffusion in PCN seems to be consistent with a power law, $\alpha_{\mathcal{C}}(\tilde{t}) \sim \tilde{t}^{-\beta}$, where in our case the characteristic exponent is $\beta \simeq 1.1$. It is worth mentioning that we achieved the very same results for PCN constructed without considering the lower filter at 4 Å, that is, by considering all contacts within 8 Å (data not shown).

Similar anomalies of functional form have been observed in the (cumulative) distribution of many exper-

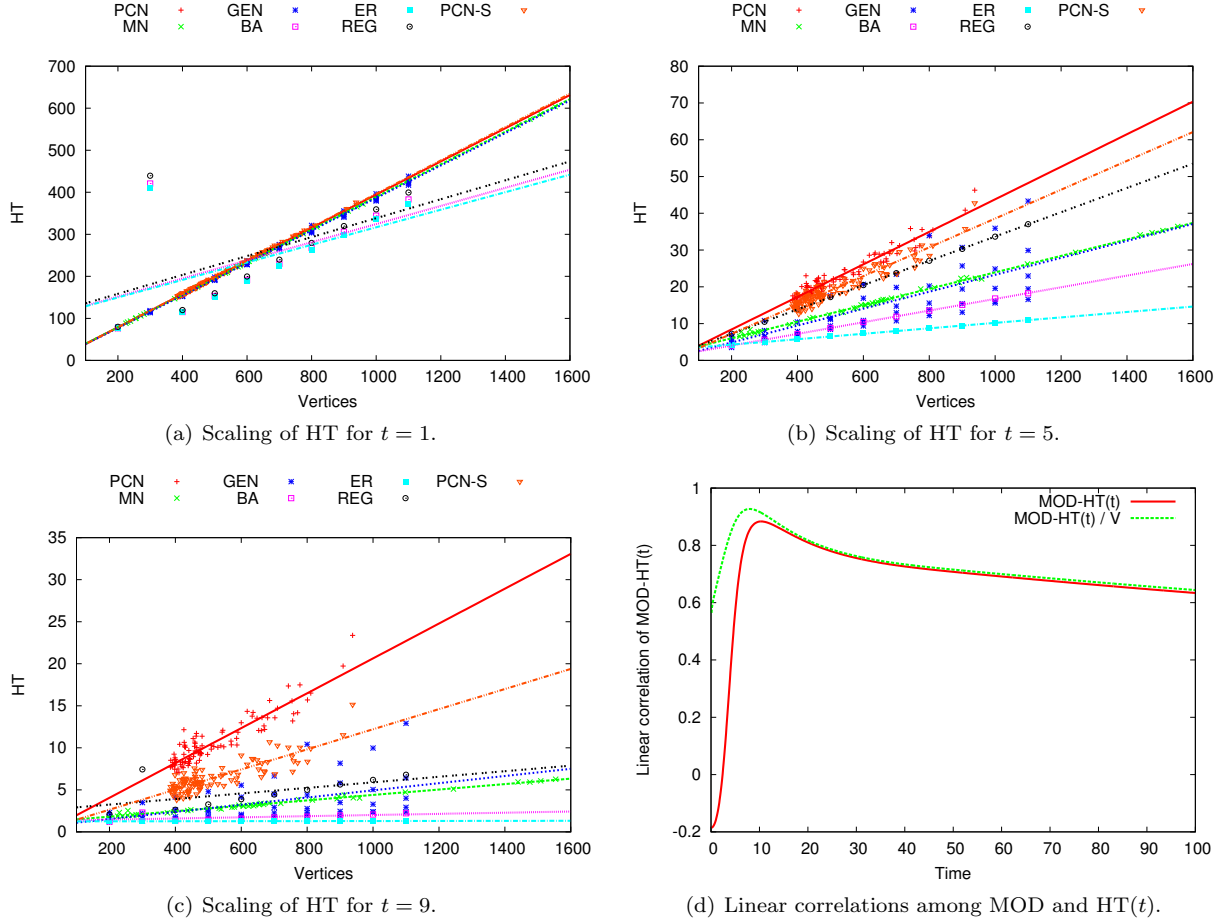


Figure 6: Scaling of HT over networks size.

imental time series, especially in those related to financial markets [45]. This phenomenon might happen when the functional form is consistent with one of the q -exponentials family, which originated in the field of non-extensive statistical mechanics [76]. In the case of PCN, this behavior is the signature of a crucial physical property of proteins, i.e., the energy flow. Energy flow in proteins mimics the transport in a three-dimensional percolation cluster [47]: energy flows readily between connected sites of the cluster and only slowly between non connected sites. This experimentally validated double regime seems to be captured by the HT decay trend shown in Fig. 7(b). We stress that this result is elaborated from the herein exploited minimalistic PCN model, so confirming the relevance of graph-based representations in protein science.

Now let us consider the results for the HC (Figs. 8(a), 8(b), and 8(c)). Those three figures depict the scaling of the HC over the network size, considering the information of the entire HM. Notably, PCN and PCN-S are the only network types showing a consistent linear scaling with the size for all time instants. Other networks are not well-described by a linear fit as the time increases. Finally, in Fig. 8(d) we show the scaling of the first HCI coefficient with the vertices (please note that for $m = 1$, Eq. 11 yields negative values). Of notable interest is the fact that PCN denote a nearly constant trend. This means that, since the HCIs are time-independent features synthetically describing the HC information, PCN denote a similar characteristic in this respect, as in fact HC scaling in Fig. 8 is consistently preserved over time.

In Fig. 9 we offer a visual representation of the heat diffusion pattern over time that is observable through the entire HM. We considered two exemplar networks of exactly the same size: the “JW0058” protein and the synthetic counterpart belonging to PCN-S that we denote here as “JW0058-SYNTH”. As discussed

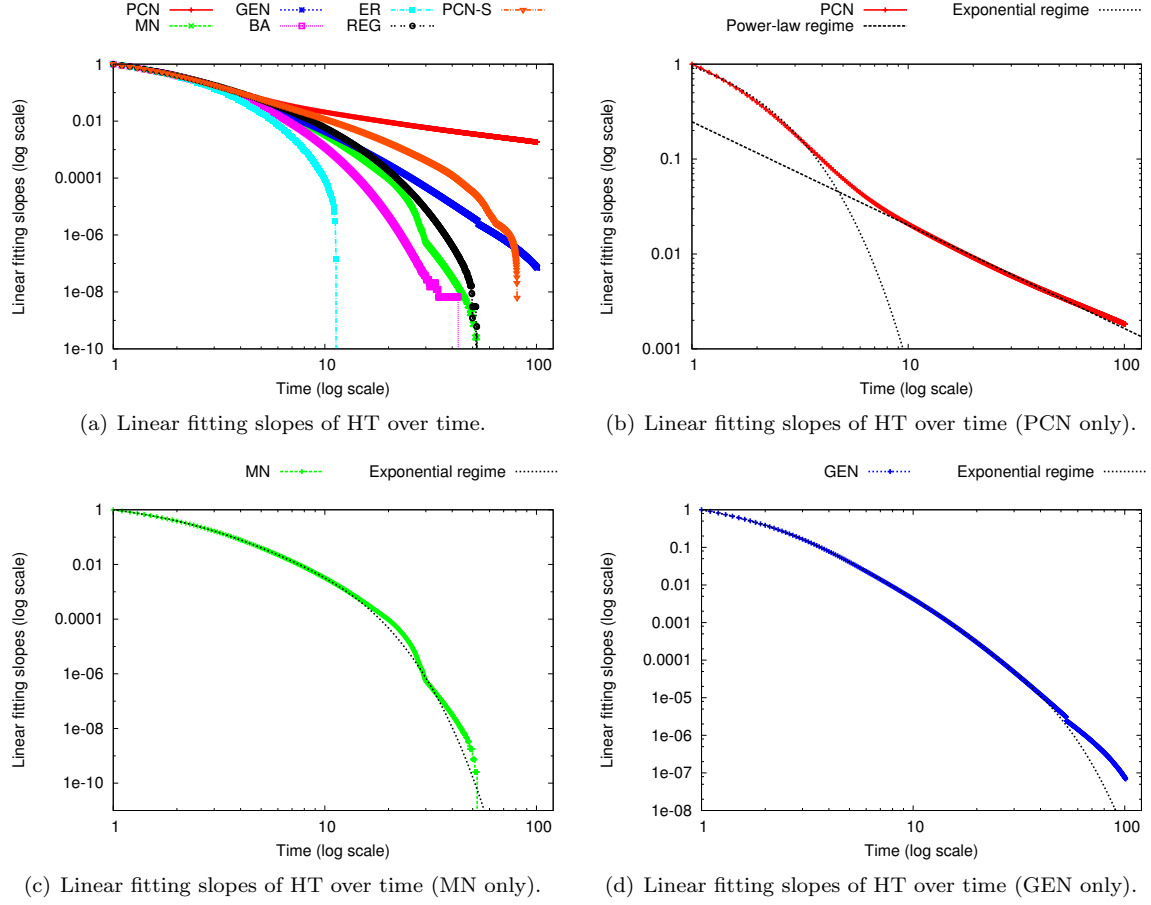


Figure 7: Scaling of HT linear best fitting slopes over time (time is sampled in 1000 equally-spaced points between 0 and 100).

before, PCN are characterized by a highly modular and fractal structure, while the considered synthetic counterpart exhibits a typical small world topology. Accordingly, by comparing the diffusion occurring on the two networks over time, it is possible to recognize significantly different patterns that were not noted in the scalings of Fig. 8. Of course, initially ($t = 1$) the heat is mostly concentrated in the vertices, which results in a very intense trace. As the time increases, the diffusion pattern for the real protein is more evident and also persistent. This is in agreement with recent laboratory experiments [47, 48], which demonstrated that diffusion in proteins proceeds slower than normal diffusion. Conversely, the diffusion for JW0058-SYNTH is in general faster since in fact the trace vanishes quickly. In graph-theoretical terms, this means that the spectral gap of PCN dominates Eq. 5 as t becomes large. This result suggests us that, in future research studies, it could be interesting to devote focused attention to the properties of the spectral density of the ensembles. Analogue results have been obtained by considering the other network types; we do not show them for the sake of brevity.

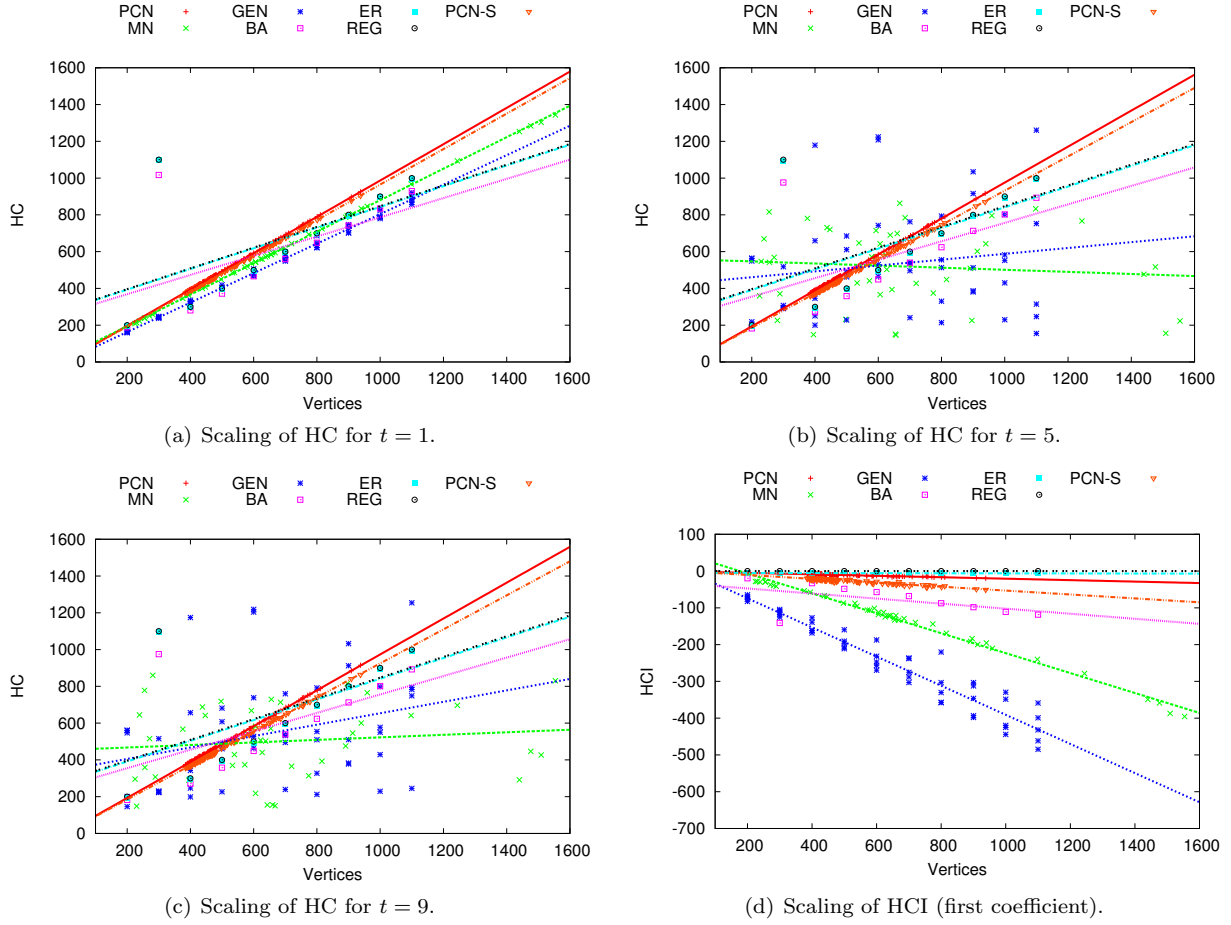
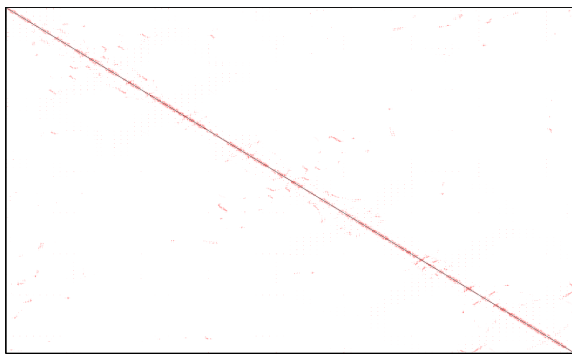
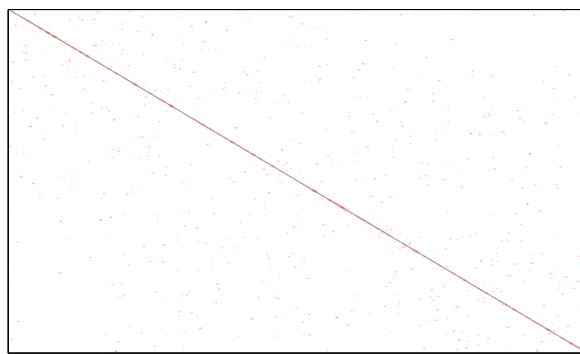


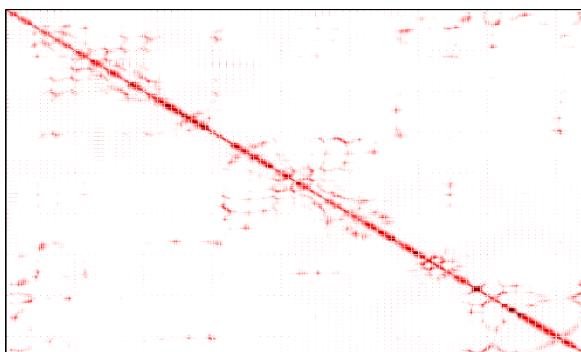
Figure 8: Scaling of HC and HCI over network size.



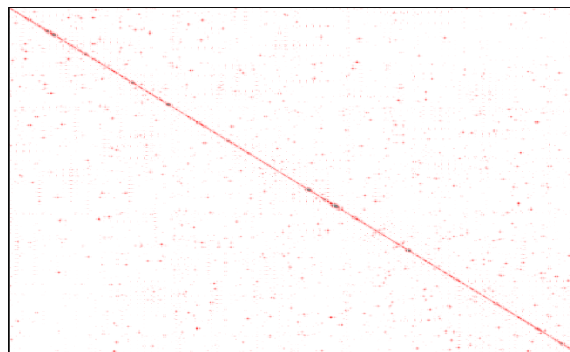
(a) HM for JW0058 at $t = 1$.



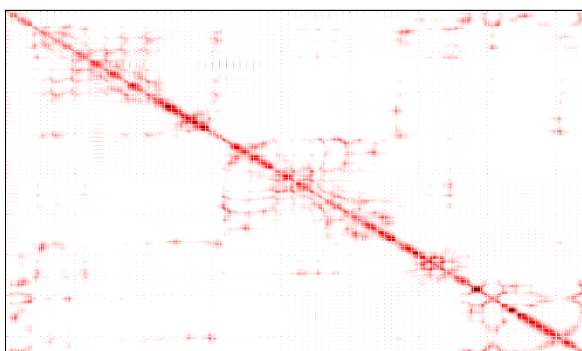
(b) HM for JW0058-SYNTH at $t = 1$.



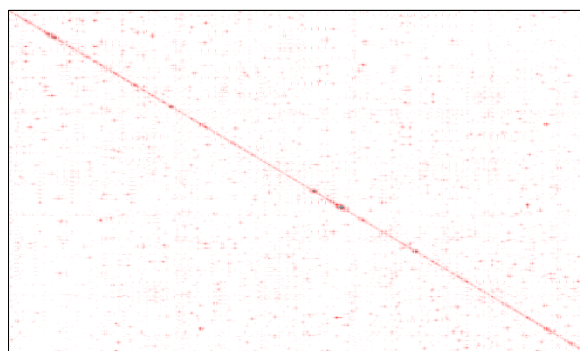
(c) HM for JW0058 at $t = 5$.



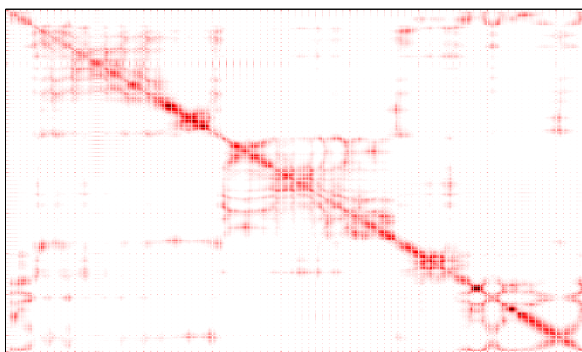
(d) HM for JW0058-SYNTH at $t = 5$.



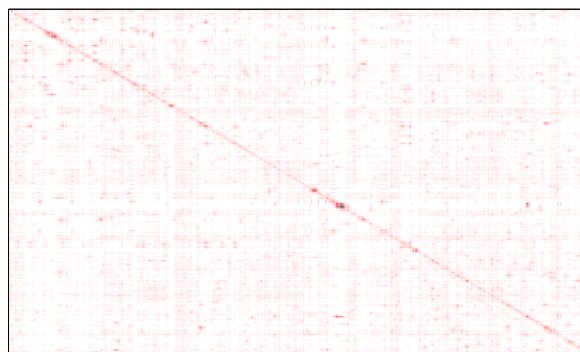
(e) HM for JW0058 at $t = 9$.



(f) HM for JW0058-SYNTH at $t = 9$.



(g) HM for JW0058 at $t = 15$.



(h) HM for JW0058-SYNTH at $t = 15$.

Figure 9: HM diffusion pattern over time for protein JW0058 and its synthetic counterpart.

4 Conclusions

In this paper we have investigated the structure of three types of complex networks: protein contact networks, metabolic networks, and gene regulatory networks, together with simulated archetypal models acting as probes. We biased the study on protein contact networks, highlighting their peculiar structure with respect to the other networks. Our analysis focused on ensemble statistics, that is, we analyzed the features elaborated by considering several instances of varying size of such networks. We considered two main network characterizations: the first one based on classical topological descriptors, while the second one exploited several invariants extracted from the discrete heat kernel. We found strong statistical agreement among those two representations, which allowed for a consistent interpretation of the results in terms of principal component analysis. Our major result was the demonstration of a double regime characterizing a (simulated) diffusion process in the considered protein contact networks. As shown by laboratory experiments, both energy flow and vibration dynamics in proteins exhibit subdiffusive properties, i.e., slower-than-normal diffusion [47]. The notable difference in the diffusion pattern between real proteins and the herein considered simulated polymers (whose contact networks have the same local structure of the corresponding real proteins), points to a peculiar mesoscopic organization of proteins going beyond the pure backbone folding. The observed correlations between MOD and HT indicates this principle in the presence of well-characterized domains. The novelty of our results is that we were able to demonstrate such a well-known property of proteins by exploiting graph-based representations and computational tools only. The fact that the observed properties emerged with no explicit reference to chemico-physical characterization of proteins, relying hence on pure topological properties, suggests the existence of general universal mesoscopic principles fulfilling the hopes expressed by Laughlin *et al.* [46].

A Heat Kernel and the Related Invariants

Let $G = (\mathcal{V}, \mathcal{E})$ be a graph (network) with $n = |\mathcal{V}|$ vertices and $m = |\mathcal{E}|$ edges. Let $\mathbf{A}^{n \times n}$ be the adjacency matrix defined as $A_{ij} = 1$ if there is an edge between vertices $v_i, v_j \in \mathcal{V}$; $A_{ij} = 0$ otherwise. Let us define the degree of a vertex v_i as $\deg(v_i) = \sum_{j=1}^n A_{ij}$. In addition, let us define \mathbf{D} as a diagonal matrix of degree: $D_{ii} = \deg(v_i)$. Let $\mathbf{L} = \mathbf{D} - \mathbf{A}$ be the Laplacian matrix of G . The normalized Laplacian matrix is given by $\hat{\mathbf{L}} = \mathbf{D}^{-1/2} \mathbf{L} \mathbf{D}^{-1/2}$, and as a consequence $\hat{\mathbf{L}}$ is symmetric and positive semi-definite; therefore it has non-negative eigenvalues only. Let us define the spectral decomposition of the Laplacian as $\hat{\mathbf{L}} = \Phi \Lambda \Phi^T$, where Λ is the diagonal matrix containing the eigenvalues arranged as $0 = \lambda_1 \leq \lambda_2 \leq \dots \leq \lambda_n \leq 2$; Φ contains the corresponding (unitary) eigenvectors as columns.

The heat equation [79] associated to the normalized Laplacian, $\hat{\mathbf{L}}$, is given by

$$\frac{\partial \mathbf{H}_t}{\partial t} = -\hat{\mathbf{L}} \mathbf{H}_t, \quad (2)$$

where \mathbf{H}_t is a doubly-stochastic $n \times n$ matrix, called heat matrix (HM), and t is the time variable. It is well-known that the solution to (2) is

$$\mathbf{H}_t = \exp(-\hat{\mathbf{L}}t), \quad (3)$$

which can be solved by exponentiating the spectrum of $\hat{\mathbf{L}}$:

$$\mathbf{H}_t = \Phi \exp(-\Lambda t) \Phi^T = \sum_{i=1}^n \exp(-\lambda_i t) \phi_i \phi_i^T. \quad (4)$$

Eq. 2 describes the diffusion of heat/information across the graph over time. In fact,

$$\mathbf{H}_t(v, u) = \sum_{i=1}^n \exp(-\lambda_i t) \phi_i(v) \phi_i(u), \quad (5)$$

where $\phi_i(v)$ is the value related to the vertex v in the i th eigenvector. It is important to note that $\mathbf{H}_t \simeq \mathbf{I} - \hat{\mathbf{L}}t$ when $t \rightarrow 0$; conversely, when t is large $\mathbf{H}_t \simeq \exp(-\lambda_2 t) \phi_2 \phi_2^T$, where ϕ_2 is the normalized Fiedler vector. This means that the large-time behavior of the diffusion depends on the global structure of the graph (e.g., its global architectural organization), while its short-time characteristics are determined by the local structure.

The heat trace (HT) of \mathbf{H}_t is given by

$$\text{HT}(t) = \text{Tr}(\mathbf{H}_t) = \sum_{i=1}^n \exp(-\lambda_i t), \quad (6)$$

which takes into account only the eigenvalues of $\hat{\mathbf{L}}$. The heat content (HC) of \mathbf{H}_t is defined by considering also the eigenvectors of $\hat{\mathbf{L}}$:

$$\text{HC}(t) = \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} \mathbf{H}_t(u, v) = \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} \sum_{i=1}^n \exp(-\lambda_i t) \phi_i(v) \phi_i(u). \quad (7)$$

Eq. 7 can be described in terms of power series expansion,

$$\text{HC}(t) = \sum_{m=0}^{\infty} q_m t^m. \quad (8)$$

By using the McLaurin series for the exponential function, we have

$$\exp(-\lambda_i t) = \sum_{m=0}^{\infty} \frac{(-\lambda_i)^m t^m}{m!}, \quad (9)$$

which substituted in Eq. 7 gives:

$$\text{HC}(t) = \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} \sum_{i=1}^n \exp(-\lambda_i t) \phi_i(v) \phi_i(u) = \sum_{m=0}^{\infty} \sum_{u \in \mathcal{V}} \sum_{v \in \mathcal{V}} \sum_{i=1}^n \phi_i(v) \phi_i(u) \frac{(-\lambda_i)^m t^m}{m!}. \quad (10)$$

The q_m coefficients in (8) are graph invariants (called heat content invariants, HCI) that can be calculated in closed-form by using Eqs. 8 and 10:

$$q_m = \sum_{i=1}^n \left(\sum_{u \in \mathcal{V}} \phi_i(u) \right)^2 \frac{(-\lambda_i)^m}{m!}. \quad (11)$$

B Ensemble Heat Trace

The HT (6) of a graph $G = (\mathcal{V}, \mathcal{E})$ with $n = |\mathcal{V}|$ can be expressed as a function of time

$$\text{HT}_G(t) = \sum_{i=1}^n \exp(-\lambda_i t) = 1 + \sum_{i=2}^n \exp(-\lambda_i t). \quad (12)$$

where λ_i are the eigenvalues of the normalized Laplacian of G . Let us define an ensemble of graphs \mathcal{C} , in which all graphs share a common characteristic spectral density. Such spectra can be synthetically described by considering the spectral density of the ensemble \mathcal{C} . Accordingly, we can consider the eigenvalues as i.i.d. random variables $\tilde{\lambda}_i$, assuming values according to the spectral density of the ensemble – except for $\tilde{\lambda}_1$ that assumes deterministically the value 0. The HT of a generic graph $G \in \mathcal{C}$, $n = |\mathcal{V}|$, can be written as:

$$\text{HT}_G(t; n) = 1 + \sum_{i=2}^n \exp(-\tilde{\lambda}_i t) = 1 + \sum_{i=2}^n \exp(-\tilde{\lambda} t). \quad (13)$$

The last step in Eq. 13 is carried out by considering that, since the $\tilde{\lambda}_i$ are assumed as i.i.d., their values can be actually expressed as n realizations of a single random variable, $\tilde{\lambda}$. For a fixed value of time t , we can define the *ensemble* HT, $\text{HT}_{\mathcal{C}}(n; t)$, as the mean HT over all graphs of the ensemble \mathcal{C} with varying size n , which is given by:

$$\text{HT}_{\mathcal{C}}(n; t) = \langle \text{HT}_G(t; n) \rangle_{\mathcal{C}} = 1 + \sum_{i=2}^n \langle \exp(-\tilde{\lambda}t) \rangle_{\mathcal{C}} = 1 + (n-1) \langle \exp(-\tilde{\lambda}t) \rangle_{\mathcal{C}}. \quad (14)$$

Hence, $\text{HT}_{\mathcal{C}}(n; t)$ can be expressed as a linear function of the graph size

$$\text{HT}_{\mathcal{C}}(n; t) = 1 - \alpha_{\mathcal{C}}(t) + \alpha_{\mathcal{C}}(t) \cdot n \simeq \alpha_{\mathcal{C}}(t) \cdot n, \quad (15)$$

where $\alpha_{\mathcal{C}}(t) = \langle \exp(-\tilde{\lambda}t) \rangle_{\mathcal{C}} \in [0, 1]$ is a time-dependent angular coefficient (slope) that is characteristic for the entire ensemble \mathcal{C} .

References

- [1] DREAM5. URL <http://wiki.c2b2.columbia.edu/dream/index.php/D5c3>.
- [2] Protein Data Bank. URL <http://www.rcsb.org/pdb/home/home.do>.
- [3] K. Anand and G. Bianconi. Entropy measures for networks: Toward an information theory of complex topologies. *Physical Review E*, 80:045102, Oct 2009. doi: 10.1103/PhysRevE.80.045102.
- [4] G. Bagler and S. Sinha. Network properties of protein structures. *Physica A: Statistical Mechanics and its Applications*, 346(1):27–33, 2005.
- [5] J. R. Banavar and A. Maritan. Physics of proteins. *Annual Review of Biophysics and Biomolecular Structure*, 36:261–280, 2007.
- [6] A. Banerji and I. Ghosh. Fractal symmetry of protein interior: what have we learned? *Cellular and Molecular Life Sciences*, 68(16):2711–2737, 2011.
- [7] A.-L. Barabási and R. Albert. Emergence of scaling in random networks. *Science*, 286(5439):509–512, 1999.
- [8] A.-L. Barabasi and Z. N. Oltvai. Network biology: understanding the cell’s functional organization. *Nature Reviews Genetics*, 5(2):101–113, 2004.
- [9] M. Barthélemy. Spatial networks. *Physics Reports*, 499(1):1–101, 2011.
- [10] L. Bartoli, P. Fariselli, and R. Casadio. The effect of backbone on the small-world properties of protein contact maps. *Physical Biology*, 4(4):L1, 2007.
- [11] A. Bashan, R. P. Bartsch, J. W. Kantelhardt, S. Havlin, and P. C. Ivanov. Network physiology reveals relations between network topology and physiological function. *Nature communications*, 3:702, 2012.
- [12] D. Ben-Avraham and S. Havlin. *Diffusion and reactions in fractals and disordered systems*. Cambridge University Press, 2000.
- [13] V. D. Blondel, J.-L. Guillaume, R. Lambiotte, and E. Lefebvre. Fast unfolding of communities in large networks. *Journal of Statistical Mechanics: Theory and Experiment*, 2008(10):P10008, 2008.
- [14] S. Boccaletti, V. Latora, Y. Moreno, M. Chavez, and D. Hwang. Complex networks: Structure and dynamics. *Physics Reports*, 424(4-5):175–308, Feb 2006. ISSN 03701573. doi: 10.1016/j.physrep.2005.10.009.
- [15] C. Böde, I. A. Kovács, M. S. Szalay, R. Palotai, T. Korcsmáros, and P. Csermely. Network analysis of protein dynamics. *Febs Letters*, 581(15):2776–2782, 2007.
- [16] E. T. Bullmore and O. Sporns. The economy of brain network organization. *Nature Reviews Neuroscience*, 13(5):336–349, 2012. doi: 10.1038/nrn3214.
- [17] L. d. F. Costa, F. A. Rodrigues, G. Travieso, and P. R. Villas Boas. Characterization of complex networks: A survey of measurements. *Advances in Physics*, 56(1):167–242, 2007.

- [18] P. Csermely, K. Singh Sandhu, E. Hazai, Z. Hoksza, H. J M Kiss, F. Miozzo, D. V. Veres, F. Piazza, and R. Nussinov. Disordered proteins and network disorder in network descriptions of protein structure, dynamics and function: hypotheses and a comprehensive review. *Current Protein and Peptide Science*, 13(1):19–33, 2012.
- [19] P. Csermely, T. Korcsmáros, H. J. M. Kiss, G. London, and R. Nussinov. Structure and dynamics of molecular networks: A novel paradigm of drug discovery: A comprehensive review. *Pharmacology & therapeutics*, 138(3):333–408, 2013.
- [20] P. Csermely, A. London, L.-Y. Wu, and B. Uzzi. Structure and dynamics of core/periphery networks. *Journal of Complex Networks*, 1(2):93–123, 2013.
- [21] M. Dehmer and A. Mowshowitz. A history of graph entropy measures. *Information Sciences*, 181(1):57–78, 2011. ISSN 0020-0255. doi: 10.1016/j.ins.2010.08.041.
- [22] M. Dehmer, L. A. J. Mueller, and F. Emmert-Streib. Quantitative network measures as biomarkers for classifying prostate cancer disease states: a systems approach to diagnostic biomarkers. *PloS one*, 8(11):e77602, 2013.
- [23] J.-C. Delvenne, S. N. Yaliraki, and M. Barahona. Stability of graph communities across time scales. *Proceedings of the National Academy of Sciences*, 107(29):12755–12760, 2010. doi: 10.1073/pnas.0903215107.
- [24] L. Di Paola, M. De Ruvo, P. Paci, D. Santoni, and A. Giuliani. Protein contact networks: an emerging paradigm in chemistry. *Chemical Reviews*, 113(3):1598–1613, 2012.
- [25] L. Di Paola, P. Paci, D. Santoni, M. De Ruvo, and A. Giuliani. Proteins as sponges: a statistical journey along protein structure organization principles. *Journal of chemical information and modeling*, 52(2):474–482, 2012.
- [26] S. N. Dorogovtsev, A. V. Goltsev, and J. F. F. Mendes. Critical phenomena in complex networks. *Reviews of Modern Physics*, 80(4):1275, 2008.
- [27] A. Duardo-Sánchez, C. R. Munteanu, P. Riera-Fernández, A. López-Díaz, A. Pazos, and H. González-Díaz. Modeling complex metabolic reactions, ecological systems, and financial and legal networks with MIANN models based on markov-wiener node descriptors. *Journal of Chemical Information and Modeling*, 54(1):16–29, 2013. doi: 10.1021/ci400280n.
- [28] M. B. Enright and D. M. Leitner. Mass fractal dimension and the compactness of proteins. *Physical Review E*, 71:011912, Jan 2005. doi: 10.1103/PhysRevE.71.011912.
- [29] F. Escolano, E. R. Hancock, and M. A. Lozano. Heat diffusion: Thermodynamic depth complexity of networks. *Physical Review E*, 85(3):036206, 2012.
- [30] E. Estrada. Universality in protein residue networks. *Biophysical Journal*, 98(5):890–900, 2010.
- [31] H. Frauenfelder and P. G. Wolynes. Biomolecules: where the physics of complexity and simplicity meet. *Physics Today*, 47(2):58–66, 1994.
- [32] L. K. Gallos, C. Song, S. Havlin, and H. A. Makse. Scaling theory of transport in complex biological networks. *Proceedings of the National Academy of Sciences*, 104(19):7746–7751, 2007. doi: 10.1073/pnas.0700250104.
- [33] L. K. Gallos, H. A. Makse, and M. Sigman. A small world of weak ties provides optimal global integration of self-similar modules in functional brain networks. *Proceedings of the National Academy of Sciences*, 109(8):2825–2830, 2012. doi: 10.1073/pnas.1106612109.
- [34] A. Giuliani, S. Filippi, and M. Bertolaso. Why network approach can promote a new way of thinking in biology. *Frontiers in genetics*, 5(83), 2014.
- [35] H. González-Díaz and P. Riera-Fernández. New markov-autocorrelation indices for re-evaluation of links in chemical and biological complex networks used in metabolomics, parasitology, neurosciences, and epidemiology. *Journal of Chemical Information and Modeling*, 52(12):3331–3340, 2012. doi: 10.1021/ci300321f.
- [36] R. Granek. Proteins as fractals: role of the hydrodynamic interaction. *Physical Review E*, 83(2):020902, 2011.
- [37] N. Guelzim, S. Bottani, P. Bourguin, and F. Képès. Topological and causal structure of the yeast transcriptional regulatory network. *Nature genetics*, 31(1):60–63, 2002.
- [38] R. Guimera, M. Sales-Pardo, and L. A. N. Amaral. Modularity from fluctuations in random graphs and complex networks. *Physical Review E*, 70(2):025101, 2004.
- [39] I. Gutman and B. Zhou. Laplacian energy of a graph. *Linear Algebra and its Applications*, 414(1):29–37, 2006.
- [40] L. Han, F. Escolano, E. R. Hancock, and R. C. Wilson. Graph characterizations from von Neumann entropy. *Pattern Recognition Letters*, 33(15):1958–1967, 2012. ISSN 0167-8655. doi: 10.1016/j.patrec.2012.03.016.

- [41] F. Höfling and T. Franosch. Anomalous transport in the crowded world of biological cells. *Reports on Progress in Physics*, 76(4):046602, 2013. doi: 10.1088/0034-4885/76/4/046602.
- [42] P. Holme and J. Saramäki. Temporal networks. *Physics Reports*, 519(3):97–125, 2012.
- [43] Y. Imry. *Introduction to mesoscopic physics*. Oxford Univ. Press, 1997.
- [44] H. Jeong, B. Tombor, R. Albert, Z. N. Oltvai, and A.-L. Barabási. The large-scale organization of metabolic networks. *Nature*, 407(6804):651–654, 2000.
- [45] J. Kwapień and S. Drożdż. Physical approach to complex systems. *Physics Reports*, 515(3):115–226, 2012. doi: 10.1016/j.physrep.2012.01.007.
- [46] R. B. Laughlin, D. Pines, J. Schmalian, B. P. Stojković, and P. Wolynes. The middle way. *Proceedings of the National Academy of Sciences*, 97(1):32–37, 2000.
- [47] D. M. Leitner. Energy Flow in Proteins. *Annual Review of Physical Chemistry*, 59(1):233–259, 2008. doi: 10.1146/annurev.physchem.59.032607.093606. PMID: 18393676.
- [48] A. Lervik, F. Bresme, S. Kjelstrup, D. Bedeaux, and J. M. Rubi. Heat transfer in protein–water interfaces. *Physical Chemistry Chemical Physics*, 12(7):1610–1617, 2010.
- [49] G. Li, D. Magana, and R. B. Dyer. Anisotropic energy flow and allosteric ligand binding in albumin. *Nature communications*, 5, 2014.
- [50] L. Livi and A. Rizzi. Graph ambiguity. *Fuzzy Sets and Systems*, 221:24–47, 2013. ISSN 0165-0114. doi: 10.1016/j.fss.2013.01.001.
- [51] L. Livi, A. Giuliani, and A. Rizzi. Toward a multilevel representation of protein molecules: comparative approaches to the aggregation/folding propensity problem. *ArXiv preprint arXiv:1407.7559*, Jul 2014.
- [52] L. Livi, A. Giuliani, and A. Sadeghian. Characterization of graphs for protein structure modeling and recognition of solubility. *arXiv preprint arXiv:1407.8033*, Jul 2014.
- [53] A. Mirshahvalad, A. V. Esquivel, L. Lizana, and M. Rosvall. Dynamics of interacting information waves in networks. *Physical Review E*, 89(1):012809, 2014.
- [54] M. Mitrović and B. Tadić. Spectral and dynamical properties in classes of sparse networks with mesoscopic inhomogeneities. *Physical Review E*, 80(2):026123, 2009.
- [55] H. Morita and M. Takano. Residue network in protein native structure belongs to the universality class of a three-dimensional critical percolation cluster. *Physical Review E*, 79:020901, Feb 2009. doi: 10.1103/PhysRevE.79.020901.
- [56] T. Nakayama, K. Yakubo, and R. L. Orbach. Dynamical properties of fractal networks: Scaling, numerical simulations, and physical realizations. *Reviews of Modern Physics*, 66(2):381, 1994.
- [57] T. Neusius, I. Daidone, I. M. Sokolov, and J. C. Smith. Subdiffusion in peptides originates from the fractal-like structure of configuration space. *Physical Review Letters*, 100(18):188103, 2008.
- [58] M. E. J. Newman. Power laws, pareto distributions and zipf’s law. *Contemporary Physics*, 46(5):323–351, 2005.
- [59] M. E. J. Newman. Modularity and community structure in networks. *Proceedings of the National Academy of Sciences*, 103(23):8577–8582, 2006.
- [60] M. E. J. Newman. *Networks: an introduction*. Oxford University Press, 2010.
- [61] V. Nicosia, M. D. Domenico, and V. Latora. Characteristic exponents of complex networks. *EPL (Europhysics Letters)*, 106(5):58005, 2014.
- [62] T. Niwa, B.-W. Ying, K. Saito, W. Jin, S. Takada, T. Ueda, and H. Taguchi. Bimodal protein solubility distribution revealed by an aggregation analysis of the entire ensemble of Escherichia coli proteins. *Proceedings of the National Academy of Sciences*, 106(11):4201–4206, 2009. doi: 10.1073/pnas.0811922106.
- [63] M. Orozco. A theoretical view of protein dynamics. *Chemical Society Reviews*, 43:5051–5066, 2014. doi: 10.1039/C3CS60474H.
- [64] A. Pinna, N. Soranzo, and A. De La Fuente. From knockouts to networks: establishing direct cause-effect relationships through graph analysis. *PloS one*, 5(10):e12912, 2010.

- [65] A. Pinna, N. Soranzo, I. Hoeschele, and A. de la Fuente. Simulating systems genetics data with sysgensim. *Bioinformatics*, 27(17):2459–2462, 2011.
- [66] S. Reuveni, R. Granek, and J. Klafter. Anomalies in the vibrational dynamics of proteins are a consequence of fractal-like structure. *Proceedings of the National Academy of Sciences*, 107(31):13696–13700, 2010.
- [67] S. Reuveni, J. Klafter, and R. Granek. Dynamic structure factor of vibrating fractals: Proteins as a case study. *Physical Review E*, 85(1):011906, 2012.
- [68] P. Riera-Fernandez, C. R. Munteanu, M. Escobar, F. Prado-Prado, R. Martín-Romalde, D. Pereira, K. Villalba, A. Duardo-Sanchez, and H. González-Díaz. New Markov–Shannon entropy models to assess connectivity quality in complex networks: from molecular to cellular pathway, parasite–host, neural, industry, and legal–social networks. *Journal of Theoretical Biology*, 293:174–188, 2012. doi: 10.1016/j.jtbi.2011.10.016.
- [69] L. Rossi, A. Torsello, E. R. Hancock, and R. C. Wilson. Characterizing graph symmetries through quantum jensen-shannon divergence. *Physical Review E*, 88(3):032806, 2013.
- [70] M. Rosvall, A. V. Esquivel, A. Lancichinetti, J. D. West, and R. Lambiotte. Memory in network flows and its effects on spreading dynamics and community detection. *Nature Communications*, 5, 2014.
- [71] A. K. Sangha and T. Keyes. Proteins fold by subdiffusion of the order parameter. *The Journal of Physical Chemistry B*, 113(48):15886–15894, 2009.
- [72] C. Song, S. Havlin, and H. A. Makse. Self-similarity of complex networks. *Nature*, 433(7024):392–395, 2005.
- [73] C. Song, S. Havlin, and H. A. Makse. Origins of fractality in the growth of complex networks. *Nature Physics*, 2(4):275–281, 2006.
- [74] M. P. H. Stumpf, T. Thorne, E. de Silva, R. Stewart, H. J. An, M. Lappe, and C. Wiuf. Estimating the size of the human interactome. *Proceedings of the National Academy of Sciences*, 105(19):6959–6964, 2008.
- [75] N. Trinajstić. *Chemical graph theory*. CRC Press, Boca Raton, FL, 1983.
- [76] C. Tsallis. I. nonextensive statistical mechanics and thermodynamics: Historical background and present status. In *Nonextensive statistical mechanics and its applications*, pages 3–98. Springer, 2001.
- [77] M. S. Vijayabaskar and S. Vishveshwara. Interaction energy based protein structure networks. *Biophysical Journal*, 99(11):3704–3715, 2010.
- [78] W. Weaver. Science and complexity. In *Facets of Systems Science*, pages 449–456. Springer, 1991.
- [79] B. Xiao, E. R. Hancock, and R. C. Wilson. Graph Characteristics from the Heat Kernel Trace. *Pattern Recognition*, 42(11):2589–2606, Nov. 2009. ISSN 0031-3203. doi: 10.1016/j.patcog.2008.12.029.
- [80] W. Yan, J. Zhou, M. Sun, J. Chen, G. Hu, and B. Shen. The construction of an amino acid network for understanding protein structure and function. *Amino Acids*, 46(6):1419–1439, 2014.
- [81] X. Yu and D. M. Leitner. Anomalous diffusion of vibrational energy in proteins. *The Journal of Chemical Physics*, 119(23):12673–12679, 2003.